ORIGINAL PAPER

Chromosomal location and gene paucity of the male specific region on papaya Y chromosome

Qingyi Yu · Shaobin Hou · Roman Hobza · F. Alex Feltus · Xiue Wang · Weiwei Jin · Rachel L. Skelton · Andrea Blas · Cornelia Lemke · Jimmy H. Saw · Paul H. Moore · Maqsudul Alam · Jiming Jiang · Andrew H. Paterson · Boris Vyskot · Ray Ming

Abstract Sex chromosomes in flowering plants evolved recently and many of them remain homomorphic, including those in papaya. We investigated the chromosomal location of papaya's small male specific region of the hermaphrodite Y (Y^h) chromosome (MSY) and its genomic features. We conducted chromosome fluorescence in situ hybridization mapping of Y^h-specific bacterial artificial chromosomes (BACs) and placed the MSY near the centromere of the papaya Y chromosome. Then we sequenced five MSY BACs to examine the genomic features of this specialized region, which resulted in the largest collection of contiguous genomic DNA sequences of a Y chromosome in flowering plants. Extreme gene paucity was observed in the papaya MSY with no functional gene identified in 715 kb MSY sequences. A high density of retroelements and local sequence duplications were detected in the MSY that is suppressed for recombination. Location of the papaya MSY near the centromere might have provided recombination suppression and fostered paucity of genes in the male specific region of the Y chromosome. Our findings provide critical information for deciphering the sex chromosomes in papaya and reference information for comparative studies of other sex chromosomes in animals and plants.

Keywords *Carica papaya* · Repetitive sequences · Segmental duplication · Sex chromosome evolution · Suppression of recombination

Communicated by Y. Van de Peer.

Q. Yu · R. L. Skelton · A. Blas · R. Ming
Hawaii Agriculture Research Center,
Aiea, HI 96701, USA

S. Hou · J. H. Saw · M. Alam
Center for Advance Studies in Genomics,
Proteomics and Bioinformatics, University of Hawaii,
Honolulu, HI 96822, USA

R. Hobza · B. Vyskot
Laboratory of Plant Development Genetics,
Institute of Biophysics, Czech Academy of Sciences,
CZ-61256 Brno, Czech Republic

F. A. Feltus · C. Lemke · A. H. Paterson
Plant Genome Mapping Laboratory,
University of Georgia, Athens, GA 30602, USA

X. Wang · W. Jin · J. Jiang
Department of Horticulture,
University of Wisconsin, Madison, WI 53706, USA

A. Blas
Department of Molecular Bioscience and Bioengineering,
University of Hawaii, Honolulu, HI 96822, USA

P. H. Moore
USDA-ARS, Pacific Basin Agricultural Research Center,
Hilo, HI 96720, USA

R. Ming (✉)
Department of Plant Biology,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA
e-mail: rming@life.uiuc.edu

Introduction

Papaya (*Carica papaya* L.) is a major fruit crop in tropical and subtropical regions worldwide. It is trioecious with all three sex types, male, female, and hermaphrodite. Sex

determination in papaya is controlled by a pair of incipient sex chromosomes differentiated by a small male-specific region of the Y chromosome (MSY) (Liu et al. 2004). The Y chromosome appears to be degenerated as the YY genotype causes embryo abortion, resulting in segregation of sex types in progenies from either female or hermaphrodite fruits. Towards understanding the sex chromosome evolution in this recently evolved system, we designed experiments to directly map the MSY on the Y chromosome and sequence long stretches of genomic DNA of this male specific region. Such information would help set strategies for identifying the sex determination genes and ultimately generate true breeding hermaphrodite varieties.

There are two slightly different Y chromosomes in papaya, one controlling males, designated as Y, and the other controlling hermaphrodites, designated as Y^h (Ming et al. 2007). Female papayas are homogametic with XX chromosomes, whereas males and hermaphrodites are heterogametic with XY and XY^h chromosomes, respectively. Storey (1976) hypothesized that males and hermaphrodites had the same genotype based on frequent sex reversals of male-to-hermaphrodite and hermaphrodite-to-male flowers. Comparison of 13 male MSY and hermaphrodite MSY DNA sequences showed that they were nearly identical, suggesting that the Y and Y^h chromosomes might have originated from the same ancestral chromosome (Liu et al. 2004). Since 32 of the 35 species in the family Caricaceae are dioecious, the divergence of these two Y chromosomes might have been the result of human selection for hermaphrodites (Storey 1976). If this were the case it would put the divergence of Y and Y^h in just thousands of years. Since hermaphrodite papaya is used for fruit and papain production in most papaya growing regions, a hermaphrodite papaya bacterial artificial chromosome (BAC) library was constructed that contains the Y^h chromosome (Ming et al. 2001), leading to the discovery and preliminary characterization of the Y^h chromosome (Ma et al. 2004; Liu et al. 2004). Severe suppression of recombination was documented by high-density genetic mapping and large scale fine mapping of the sex determination gene; reduced gene density and increased repetitive sequences relative to the genome wide averages were observed by sequencing a small number of subclones from selected MSY BACs (Ma et al. 2004; Liu et al. 2004). However, the chromosomal location and genomic features from long contiguous sequences (i.e. BACs) are unknown prior to this report.

The origin of sex chromosomes from a pair of autosomes has been confirmed by comparative genomic analyses between mammals and chicken (Smith and Sinclair 2004), among species of *Drosophila* (Carvalho and Clark 2005) and *Silene* (Filatov 2005), and by the recently discovered incipient Y chromosomes in papaya and stickleback fish (Liu et al. 2004; Peichel et al. 2004). It has been postulated

that sex chromosomes evolved from linked male and female-sterile mutations followed by suppression of recombination at the sex determination locus and that this process led to the degeneration of the Y chromosomes (Charlesworth and Charlesworth 1978; Charlesworth 1991). Suppression of recombination, which is a characteristic of sex chromosomes (Darlington 1958; Westergaard 1958; Ohno 1967), could have been caused by a variety of mechanisms, including chromosomal rearrangements, DNA methylation, and accumulation of repetitive sequences (Charlesworth et al. 1986; Jaarola et al. 1998; Maloisel and Rossignol 1998). Sex chromosomes have evolved independently, even within the same genus, numerous times in animals and more recently in flowering plants (Desfeux et al. 1996; Charlesworth 2002; Carvalho and Clark 2005).

Once suppression of recombination occurs, the MSY would accumulate mutations, transposable element insertions, and local duplications, inversions, and translocations that would lead to its degeneration. The male specific region of the highly degenerated human Y chromosome contains 78 protein coding genes, including two X-transposed, 16 X-degenerated, and 60 ampliconic genes (Skaletsky et al. 2003). Ancestral genes have degenerated or been lost at a rate of 3–4 genes per million years (my) from the human Y chromosome, from an ancestral set presumed to be similar to the 1098 genes remaining on the human X chromosome (Ross et al. 2005). The medaka fish Y chromosome evolved from translocation of a duplicated fragment from linkage group (LG) 9 that contained *dmrt1a* to LG1 (Nanda et al. 2002). The medaka sex determination gene *dmrt1bY* evolved from *dmrt1a*, whose ortholog *dmrt1* is a candidate downstream sex determination gene in mammals (Raymond et al. 1999; Matsuda et al. 2002). The ancestral 43 kb sequence on LG9 contains four functional genes and one pseudogene, whereas the 258 kb MSY of medaka contains only one functional gene, which is the sex determination gene *dmrt1bY* (Kondo et al. 2006).

Genetic mapping of the papaya MSY placed it in the middle of LG1 (Ma et al. 2004). Survey sequencing of MSY BACs indicated a reduced gene density when compared to the genome wide average (Liu et al. 2004). The objectives of the present investigation are to map the MSY BACs on the Y chromosome through chromosome in situ hybridization and to assess the genomic features of the MSY by complete sequencing of selected BACs.

Materials and methods

Plant materials

Leaf tissues from female and hermaphrodite plants of the Hawaiian gynodioecious papaya cultivars Kapoho and

SunUp, plus male and female plants of the Australian dioecious variety AU9, were used for PCR. The roots, leaves, and young flower buds from female and hermaphrodite plants of SunUp and male plants of AU9 were used for RT-PCR. Young flower buds containing anthers at various stages of meiosis of the hermaphrodite cv. SunUp were used for chromosome preparations.

Sequencing MSY BACs and sequence assembly

The papaya BAC clones were sequenced using the shotgun approach with at least 10X coverage. BAC DNAs were randomly sheared using Hydroshear (Genomic Solutions, Ann Arbor, MI, USA) to generate ∼3 kb insert fragments. The sheared fragments were size-selected on an agarose gel, purified, end-repaired, and ligated to the pUC118 vector (Takara, NY, USA). DNAs from the 3-kb libraries were cycle-sequenced with ABI BigDye Terminator v3.1 and analyzed on a 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

Phred/Phrap/Consed and CAP3 packages were used for sequence assembly. Gaps in assembly and regions of low-quality were resolved by resequencing subclones identified by Autofinish, sequencing PCR products, and/or additional random subclone sequencing. All BAC clones were manually examined for signs of misassembly. Suspect regions were clarified either by ambiguous read removal, PCR amplification and sequencing, and/or alignment with a neighboring BAC. A BAC was not considered complete until all inconsistent read pairs were resolved and Consed reported an error rate of less than 1/10,000 bases.

Indentification of repeats and transcription units and comparative sequence analyses

Interspersed repeats and low-complexity regions were identified and masked using RepeatMasker (http://www.repeatmasker.genome.washington.edu) and the *Arabidopsis* repeat database. Potential transcripts were identified using GeneScan (http://www.genes.mit.edu/GENSCAN.html). The predicted transcripts were tested by RT-PCR. Genome comparison browser (http://www.sun1.softberry.com) was used for comparative sequence analysis.

RT-PCR

Total RNA was extracted from young flower buds (0.4–0.7 cm, before meiosis stage), root, and leaf tissues. One microgram of total RNA was treated with RNase-free DNase (Promega, Madison, WI, USA) and then used for cDNA synthesis using RETROscript kit (Ambion, CA, USA). The synthesized cDNAs served as templates for RT-PCR.

Preparation of chromosomes for fluorescence in situ hybridization (FISH)

Young hermaphrodite flower buds containing anthers at various stages of meiosis were fixed in 3:1 100% ethanol: glacial acetic acid and kept at 4°C until analysis. Microsporocytes at appropriate stages of prophase I were squashed in 50% acetic acid. Slides were pretreated according to the methods of Heslop-Harrison (1998) and kept at −20°C until FISH was performed.

Fluorescence in situ hybridization

Probe preparation and signal detection for FISH was as described previously (Jackson et al. 1998).

Results

Chromosome placement of the MSY

We previously demonstrated that severe suppression of recombination occurred in the MSY of the papaya Y^h chromosome (Ma et al. 2004; Liu et al. 2004). To physically locate the MSY on the Y^h chromosome, we directly hybridized two confirmed MSY BACs, 54H01 and 76M08, on interphase, prometaphase, metaphase, and anaphase chromosomes (Fig. 1). Both BACs located on or near the centromere and the conserved sequences showed weaker signals at the centromeres of other chromosomes (Fig. 1a and c). BAC 54H01 showed a particularly strong signal on the Y^h chromosome. Hybridization of 54H01 to the apex of the V-shaped anaphase chromosomes was suggestive that the MSY is near the centromere (Fig. 1b). The chromosome showing the second strongest signal is postulated to be the X chromosome (Fig. 1b), which, despite sequence divergence from the Y^h, shares more similarity with the MSY than with other chromosomes. BAC 76M08 hybridized strongly to the pair of Y^h chromosomes in a tetraploid cell arising from endoreduplication in flower buds. In this case, the signal from the X chromosome was not distinguishable from that of autosomes (Fig. 1c), suggesting more extensive sequence divergence between the X and Y^h chromosomes in this region. The paired images showed MSY BACs to be located in the more intensely stained regions (particularly the pair in Fig. 1e and h), reinforcing the notion that the MSY might be located near the centromere of the Y^h chromosome (Fig. 1d–i). The X and Y sequence divergence was further demonstrated by simultaneous pachytene FISH mapping of BAC 76M08 and the neighboring BAC 79C23, each with a non-MSY BAC on the same slide, showing a strong signal on a single chromosome
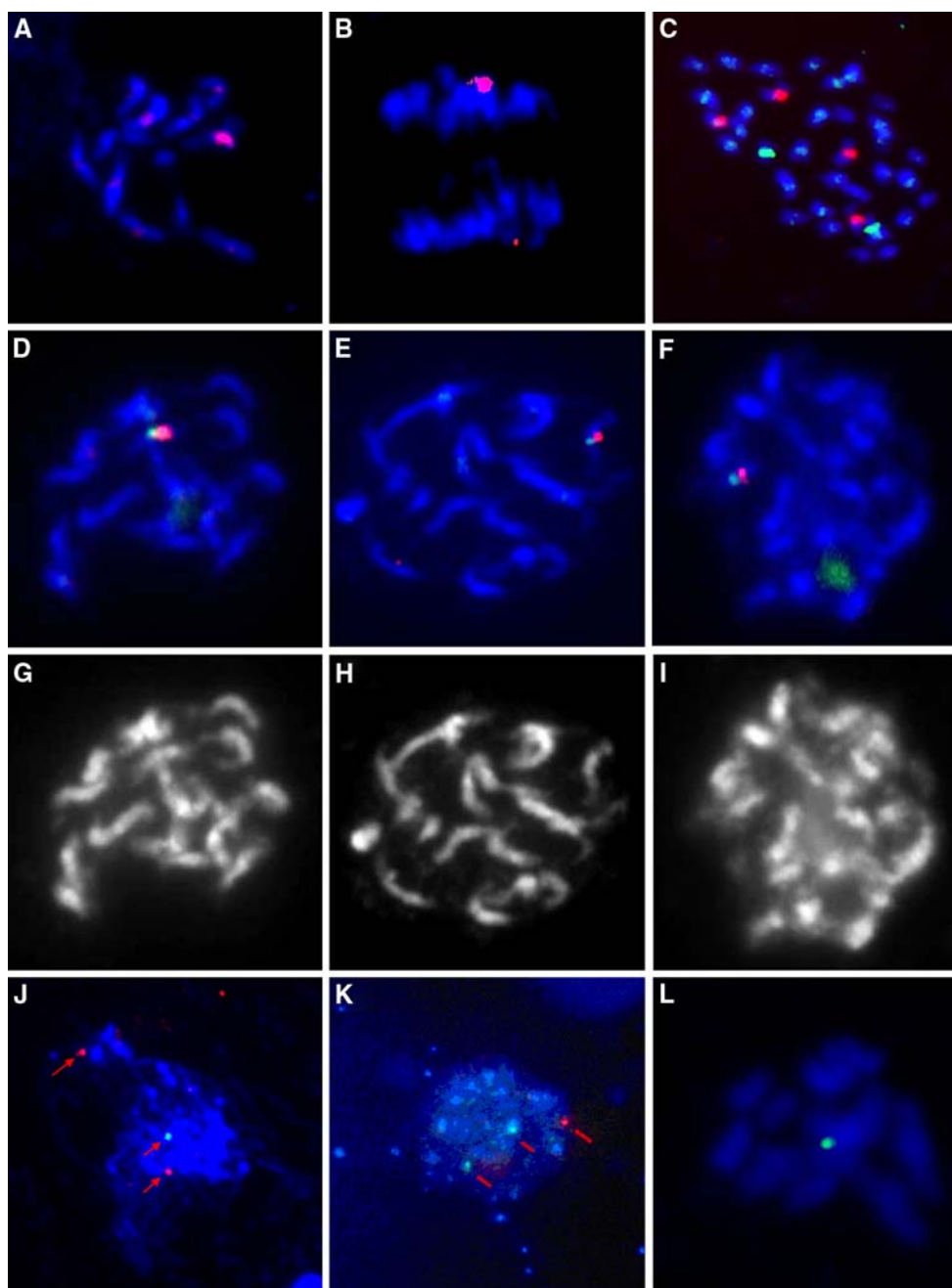
Fig. 1 FISH mapping of MSY BACs on papaya chromosomes at different stages of the cell cycle. **a** FISH mapping of MSY BAC clone 54H01 (188 kb) on papaya prometaphase chromosomes with a strong signal on the Y^h, weaker signal on the X, and faint signals on centromeres of some chromosomes. **b** MSY BAC 54H01 on anaphase chromosomes with strong signal on the Y^h and weaker signal on the X chromosome at the bottom of the V shape. **c** MSY BAC clone 76M08 (green) and *Arabidopsis thaliana* BAC F16F17 (red) from chromosome 5 containing histone-like transcription factor (CBF/NF-Y) on metaphase chromosomes of tetraploid cells with strong signals on two Y^h chromosomes and cross hybridization on centromeres of other chromosomes. **d, e** MSY BAC clones 54H01 (green) and 76M08 (red) on papaya prometaphase chromosomes. **f** MSY BAC clones 54H01 (green) and 94E22 (red) on papaya prometaphase chromosomes. **g–i** The corresponding DAPI-stained chromosomes to **d–f, j** On interphase nuclei, the MSY BAC 76M08 (green) showed a strong signal on one chromosome, while the non-MSY BAC 08J04 (red) showed signals on one pair of chromosomes. **k** MSY 79C23 (red) showed only a single FISH site, while previously identified BAC 71E16 (green) showed two sites on a different pair of interphase chromosomes, suggesting that this BAC was not on the Y^h chromosome. **l** MSY BAC 76M08 (green) on papaya prometaphase chromosomes showing signal on only one chromosome

(Fig. 1j and k) and by the mapping of BAC 76M08 alone on a metaphase chromosome (Fig. 1l). These two MSY BACs, 76M08 and 79C23, did not hybridize to their X chromosome counterparts. The non-MSY BACs, 08J04 and 71E16, showed strong signals on two homologous chromosomes (Fig. 1j and k).

Sequence analyses and characterization of five MSY BACs

To elucidate the genomic features of the MSY, we sequenced five mapped MSY BACs, 54H01, 76M08, 42B05, 41F24, and 94E22. The physical map positions of these 5 BACs on the current MSY map are shown in supplemental Fig. 1. Sequence analysis showed that none of these BACs contain known centromere-specific sequences, but they did contain numerous gypsy retroelements, which are a typical feature of the pericentromeric region of plant chromosomes (i.e. abundant gypsy and few copia retroelements). The total sequences of 714,880 bp from these five BACs contained 140,834 bp (19.7%) of known repetitive sequences based on a RepeatMasker search. The repetitive sequences included 115 gypsy retroelements totaling 110,693 bp, two copia retroelements totaling 2,138 bp, 85 simple repeats totaling 4862 bp, 457 low complexity repeats totaling 22,805 bp, one DNA transposon of 251 bp, and one small RNA of 74 bp (Fig. 2; Table 1).

Detailed sequence comparison among these five BACs revealed numerous small scale duplication events, some of which have resulted from MSY sequence divergence, following large scale duplication events (Supplemental Table 1). An ancient inverted duplication appears to have occurred in the region covered by MSY BAC 54H01, followed by a recent direct duplication event involving BACs 54H01 and 42B05 as shown by direct sequence alignment between these two BACs (Fig. 3). The recent duplication spans at least 90 kb of the BAC 42B05 with highly conserved and co-linear sequence alignment. The ancient inverted duplication is seen by sparse sequence conservation and DNA sequence expansion in the region covered by 54H01. These two duplication events were also obvious in the dot plot of gapless sequence alignment generated by BlastZ after masking the repetitive sequences (Supplemental Fig. 2).

Comparison of these five MSY BAC sequences to the entire *Arabidopsis* genome revealed matching sequences from three of them, 54H01, 76M08, and 41F24. Most matches occurred in the centromeric and pericentrimeric regions (Supplemental Fig. 3). A cluster of four tandem 79 bp repeats within 1603 bp on BAC 76M08 matched 5 S ribosomal RNA genes in the *Arabidopsis* centromeric region with 97% DNA sequence identity. A fragment of 410 bp of BAC 54H01 matched chloroplast DNA located in the *Arabidopsis* pericentromeric region with 86% DNA sequence identity.

Mining for functional genes from five MSY BACs

DNA sequences of the five MSY BACs were used to search for functional genes in the 8,571 papaya unigene set derived from flower buds (Yu et al. unpublished data). There was not a single match. A GenBank non-redundant protein database search revealed no matches to any known genes on these five BACs. After masking the repetitive

Fig. 2 Genomic structure of five MSY BACs. Distributions of repetitive sequences on the BACs are shown as vertical bars. Intrachromosomal segmental duplications of >3 kb and >80% sequence identity are shown by the same color. Orientation of the duplications is indicated by arrows. Specific sequence duplications occur across several BACs. Details of location, orientation, and extent of segmental duplications are listed in Supplemental Table 1. The chloroplast genomic DNA insertions are shown as thick green arrows
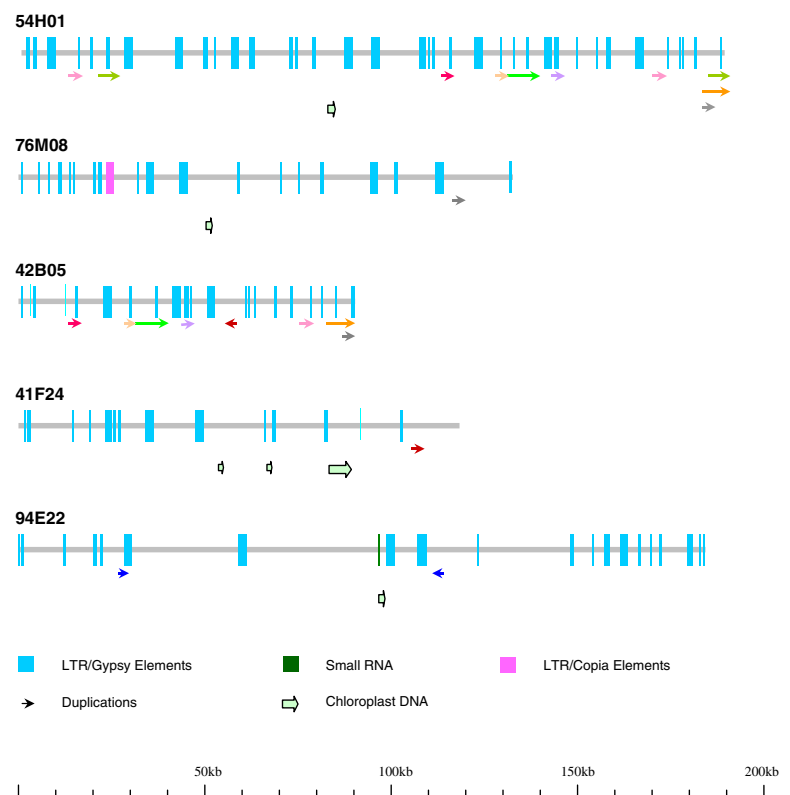
Table 1 Composition of repetitive sequences on five MSY BACs

Types of repetitive sequences	76M08 Length occupied (bp)	76M08 Percentage of sequence (%)	94E22 Length occupied (bp)	94E22 Percentage of sequence (%)	42B05 Length occupied (bp)	42B05 Percentage of sequence (%)	41F24 Length occupied (bp)	41F24 Percentage of sequence (%)	54H01 Length occupied (bp)	54H01 Percentage of sequence (%)
Retroelements	18870	14.22	22833	12.38	15782	17.42	11891	10.03	43455	23.03
SINEs	0	0	0	0	0	0	0	0	0	0
LINEs	0	0	0	0	0	0	0	0	0	0
LTR elements	18870	14.22	22833	12.38	15782	17.42	11891	10.03	43455	23.03
Ty1/Copia	2138	1.61	0	0	0	0	0	0	0	0
Gypsy/DIRS1	16732	12.61	22833	12.38	15782	17.42	11891	10.03	43455	23.03
DNA transposons	251	0.19	0	0	0	0	0	0	0	0
Total interspersed repeats	19121	14.41	22833	12.38	15782	17.42	11891	10.03	43455	23.03
Small RNA	0	0	74	0.04	0	0	0	0	0	0
Simple repeats	564	0.42	3241	1.76	172	0.19	403	0.34	482	0.26
Low complexity	2802	2.11	8769	4.76	2968	3.28	3878	3.27	4388	2.326
Bases masked	22487	16.94	34917	18.94	18922	20.89	16183	13.65	48325	25.61
Total length	132736 bp		184362 bp		90579 bp		118541 bp		188662 bp	

Fig. 3 DNA sequence comparison between MSY BACs 54H01 and 42B05. The BAC 54H01 is shown as a green horizontal line and BAC 42B05 is shown as a blue horizontal line. The location and direction of each block of duplication are indicated by a pair of triangles. The large block of duplication started from the second half of BAC 54H01 and covered the entire BAC 42B05 about 500 kb apart. The size of this duplication is likely larger than 90 kb as detected by DNA markers

sequences, 26 potential genes were predicted using Gene-Scan software. Fifty-two primer pairs, designed from the predicted genes, were used in RT-PCR using total RNA from three different tissues to verify gene expression. Nineteen of the 52 primer pairs amplified cDNA fragments, while the other 33 pairs failed to produce a PCR product. Twelve of the 19 successful amplifications produced fragments of the predicted size and seven produced fragments of different sizes. Sequencing each of the 19 fragments showed that none of them matched the DNA sequences of the MSY BACs.

Discussion

For a primitive sex chromosome system with a small MSY, the information on the chromosomal location and their genomic features is crucial for understanding the mechanism and selection forces led to the rise of the sex chromosomes. It also helps to design strategies for functional analysis of the MSY and the ultimate identification of the sex determination genes. In the study presented here we have mapped the papaya MSY near the centromere. Detailed analyses of five MSY BACs revealed extremely

low gene density, suggesting that the MSY originated from a gene poor region. The chromosomal origin of the sex chromosomes in papaya has two implications: (1) it might be universal that sex chromosomes originate from a region restricted for recombination; (2) there are diverse mechanisms to provide the initial suppression of recombination that cause the rise of sex chromosomes in addition to inversions, deletions/duplications, and translocations. The collection of long contiguous genomic sequences of a Y chromosome allows a case study of the process of Y chromosome degeneration. Based on the chromosomal location of the MSY and its gene poor nature, we expect to encounter potential heterochromatic islands on the MSY during our quest to map and sequence this genomic region, and we plan to conduct initial functional analysis of the MSY by RT-PCR and quantitative RT-PCR because of the limited number of predicted genes.

The paucity of genes uncovered in the five MSY BACs that we sequenced is quite rare, with the possible exception of degenerated Y chromosomes and to a lesser extent, chromosome centromeric regions. The papaya MSY could possibly lie within a pericentromeric region, which is recognized as usually gene poor. However, the papaya MSY was not completely devoid of genes. Four genes were identified from the two MSY BACs 95B12 and 85B24 (Yu et al. submitted). The gene density of these fully characterized BACs was 1 gene/257 kb, which contrasts with an average gene density of 1 gene/10.6 kb in the papaya genome based on an estimate from BAC end sequences (Lai et al. 2006). As an example of sequenced gene poor regions, the rice centromere CEN8 contains 1 gene/46.9 kb in the 750 kb CENH3 binding domain (A) (Yan et al. 2005); however, it should be noted that this centromere evolved recently, only 8 million years ago (mya). As a second example, the 23 Mb euchromatic human Y chromosome sequence contains 78 protein-coding genes at an average gene density of 1 gene/294 kb. The gene density in the papaya MSY region is higher than in the human Y chromosome. The human Y chromosome evolved from a common ancestral autosome shared with birds about 310 mya (Waters et al. 2007), whereas the papaya Y chromosome is calculated to be about 2 to 3 mya based on sequence divergence of X and Y gene pairs (Yu et al. submitted). The low gene density of the papaya MSY region in spite of its young age could be a consequence of its origin in a gene-poor pericentromeric region.

Local segmental duplications, transposable elements, and plastid DNA insertions, have all contributed to divergence of the papaya Y chromosome from its X chromosome counterparts. Based on the numerous small tandem duplications on 54H01 and their matching sequences on 42B05 (Fig. 3), it appears that a duplication within the MSY BAC 54H01 preceded the duplication between BACs

54H01 and 42B05. The possibility of these two duplication events is supported by the different degree of the interspersed sequences caused by insertions and deletions during the process of Y chromosome evolution (Fig. 3).

After genetic and physical mapping placed the MSY of papaya near the middle of LG 1 (Ma et al. 2004; Liu et al. 2004), we postulated that this region might be close to the centromere. Chromosome FISH mapping of MSY BACs validated this notion. The lack of centromeric-type sequences from the MSY BACs reported here support the possibility that the MSY is on only one side of the centromere and does not include it. This pericentromeric region might have provided initial reduction of recombination at the sex determination locus. Subsequent chromosomal rearrangements such as duplications and inversions could have reinforced the suppression of recombination and extended the non-recombining region. Precedence for such a possibility is seen with the *Arabidopsis* centromere 1 (CEN1) that showed 53-fold reduction in recombination frequency in the left pericentromere, a 10-fold reduction in the right pericentromere, and 200-fold reduction in the centromeric core (Haupt et al. 2001). The papaya MSY showed no recombination within about a 1 Mb region tested by sex-linked male-specific markers W11 and T12 on 1481 F_2 and F_3 plants (Liu et al. 2004). This lack of recombination is a further indication of divergence of the MSY sequence from an ancestral autosomal sequence that was shared with the X chromosome, as opposed to the original pericentromeric sequence that would recombine at a very low rate.

It should be noted that the origin of the MSY on young Y chromosomes of papaya, medaka, and sticklebacks appears to be at chromosomal regions known for restricted recombination. The medaka Y chromosome originated from insertion of a duplicated fragment on LG 9 and this inserted genomic fragment had no homologous sequence with which to pair and recombine (Kondo et al. 2004). The MSY of sticklebacks is in the telomeric region of the Y chromosome (Peichel et al. 2004), whereas the chromosomal origin of the MSY in papaya we report here appeared to be in the pericentromeric region.

The location of the papaya Y chromosome and its abundance of gypsy-type retroelements are consistent with the prevalence of gypsy-type retroelements in the centromeric and pericentromeric regions in cereals and *Arabidopsis* (Kurata et al. 2002; Jiang et al. 2003; Nagaki et al. 2003). The 13.7–25.6% masked repetitive sequences on each of the papaya MSY BACs using an *Arabidopsis* repeat database are likely underestimates of the level of repetitive sequences. For example, when the 258 kb MSY sequences of medaka were searched using RepeatMasker2 program against the *fugu* repeat database, only 4.2% of the medaka sequences were identified as repetitive sequences. However, when using a database consisting of medaka, *fugu*, and

zebrafish repeat sequences, 53.2% of the medaka MSY sequences were classified as repetitive (Kondo et al. 2006). Six of the ten most abundant papaya-specific repeats shared homology with papaya MSY sequences (Lai et al. 2006), further suggesting that a significant fraction of the repeat sequences on the papaya MSY were not masked by the *Arabidopsis* repeat database. Currently, there is no papaya specific repeat database. More detailed MSY sequence analysis will be carried out when such a repeat database is available.

The key event for evolution of sex chromosomes is the suppression of recombination at the sex determination locus. Male and female sterile mutations on nuclear DNA could occur at a chromosomal location where the mechanism of suppression of recombination was already in place, as shown in the telomeric location of the MSY in a nascent Y chromosome in stickleback fish (Peichel et al. 2004). There is only one pseudo-autosomal region on the q arm of the *Silene* Y chromosome (Di Stilio 1998; Lengerova 2003), and it appears that the suppression of recombination started from the p arm (without PAR) and spread to the q arm of the Y chromosome (Lengerova et al. 2003; Nicolas et al. 2005; Zluvova et al. 2005). Initiation of the papaya MSY near the centromere might have placed it in a region with reduced genetic recombination and allowed sequence divergence in an already gene-poor region, facilitating the evolution of a recently evolved Y chromosome.

References

Carvalho AB, Clark AG (2005) Y chromosome of *D. pseudoobscura* is not homologous to the ancestral *Drosophila* Y. Science 307:108–110

Charlesworth B (1991) The evolution of sex chromosomes. Science 251:1030–1033

Charlesworth D (2002) Plant sex determination and sex chromosomes. Heredity 88:94–101

Charlesworth B, Charlesworth D (1978) A model for the evolution of dioecy and gynodioecy. Am Nat 112:975–997

Darlington CD (1958) The evolution of genetic systems. Basic Books Inc., New York

Desfeux C, Maurice S, Henry J-P, Lejeune B, Gouyon P-H (1996) Evolution of reproductive systems in the genus *Silene*. Proc R Soc Lond B 263:409–414

Di Stilio VS, Kesseli RV, Mulcahy DL (1998) A pseudoautosomal random amplified polymorphic DNA marker for the sex chromosomes of *Silene dioica*. Genetics 148:2057–2062

Filatov DA (2005) Evolutionary history of *Silene latifolia* sex chromosomes revealed by genetic mapping of four genes. Genetics 170:975–979

Haupt W, Fischer TC, Winderl S, Fransz P, Torres-Ruiz RA (2001) The CENTROMERE 1 (CEN1) region of *Arabidopsis* thaliana: architecture and functional impact of chromatin. Plant J 27:285–296

Heslop-Harrison JS (1998) Cytogenetic analysis of Arabidopsis. In: Martinez-Zapater JM, Salinas J (Eds) Arabidopsis protocols. Humana Press, Totowa, pp. 119–127

Jackson SA, Wang M, Goodman HM, Jiang J (1998) Fiber-FISH analysis of repetitive DNA elements in *Arabidopsis thaliana*. Genome 41:566–572

Jaarola M, Martin RH, Ashley T (1998) Direct evidence for suppression of recombination with two pericentric inversions in humans: a new sperm-FISH technique. Am J Hum Genet 63:218–224

Jiang J, Birchler JA, Parrott WA, Dawe RK (2003) A molecular views of plant centromeres. Trends Plant Sci 12:570–575

Kondo M, Hornung U, Nanda I, Imai S, Sasaki T, Shimizu A, Asakawa S, Hori H, Schmid M, Shimizu N, Schartl M (2006) Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. Genome Res 16:815–826

Kondo M, Nanda I, Hornung U, Schmid M, Schartl M (2004) Evolutionary origin of the medaka Y chromosome. Curr Biol 14:1664–1669

Kurata N, Nonomura K-I, Harushima Y (2002) Rice genome organization: the centromere and genome interactions. Ann Bot 90:427–435

Lai CWJ, Yu Q, Hou S, Skelton RL, Jones MR, Lewis KLT, Murray J, Eustice M, Guan P, Agbayani R, Moore PH, Ming R, Presting GG (2006) Analysis of papaya BAC end sequences reveals first insights into the organization of a fruit tree genome. Mol Genet Genomics 276:1–12

Lengerova M, Moore RC, Grant SR, Vyskot B (2003) The sex chromosomes of *Silene latifolia* revisited and revised. Genetics 165:935–938

Liu Z, Moore PH, Ma H, Ackerman CM, Ragiba M, Yu Q, Pearl HM, Kim MS, Charlton JW, Stiles JI, Zee FT, Paterson AH, Ming R (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature 427:348–352

Ma H, Moore PH, Liu Z, Kim MS, Yu Q, Fitch MMM, Sekioka T, Paterson AH, Ming R (2004) High-density linkage mapping revealed suppression of recombination at the sex determination locus in papaya. Genetics 166:419–436

Maloisel L, Rossignol J-L (1998) Suppression of crossing-over by DNA methylation in *Ascobolus*. Genes Dev 12:1381–1389

Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, Morrey CE, Shibata N, Asakawa S, Shimizu N, Hori H, Hamaquchi S, Sakaizumi M (2002) *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. Nature 417:559–563

Ming R, Q Yu, Moore PH (2007) Sex determination in papaya. Semin Cell Dev Biol (in press)

Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jaing J (2003) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics 163:1221–1225

Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, Shimizu A, Shan Z, Haaf T, Shimizu N, Shima A, Schmid M, Schartl M (2002) A duplicated copy of *DMRT1* in the sex determining region of the Y chromosome of the medaka, *Oryzias latipes*. Proc Natl Acad Sci USA 99:11778–11783

Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D, Neqrutiu I, Charlesworth D, Moneqer F (2005) A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. PLoS Biol 3:47–56

Ohno S (1967) Sex chromosomes and sex linked genes. Springer, Berlin, pp 5–13

Peichel CL, Ross JA, Matson CK, Dickson M, Grimwood J, Schmutz J, Myers RM, Mori S, Schluter D, Kingsley DM (2004) The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. Curr Biol 14:1416–1424

Raymond CS, Parker ED, Kettlewell JR, Brown LG, Page DC, Kusz K, Jaruzelska J, Reinberg Y, Flejter WL, Bardwell VJ, Hirsch B, Zarkower D (1999) A region of human chromosome 9 p required for testis development contains two genes related to known sexual regulators. Hum Mol Genet 8:989–996

Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP et al (2005) The DNA sequence of the human X chromosome. Nature 434:325–337

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L et al (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837

Smith CA, Sinclair AH (2004) Sex determination: insights from the chicken. BioEssays 26:120–132

Storey WB (1976) Papaya. In: Simmonds NW (Ed) Evolution of crop plants. Longman, London, pp 21–24

Waters PD, Wallis MC, Graves, JAM (2007) Mammalian sex-origin and evolution of the Y chromosome and SRY. Semin Cell Dev Biol. doi: 10.1016/j.semcdb.2007.02.007

Westergaard M (1958) The mechanism of sex determination in flowering plants. Adv Genet 9:217–281

Yan H, Jin W, Nagaki K, Tian S, Ouyang S, Buell CR, Talbert PB, Henikoff S, Jiang J (2005) Transcription and histone modifications in the recombination-free region spanning a rice centromere. Plant Cell 17:3227–3238

Zluvova J, Janousek B, Negrutiu I, Vyskot B (2005) Comparison of the X and Y chromosome organization in *Silene latifolia*. Genetics 170:1431–1434